HIPOTĒŽU PĀRBAUDE UN STATISTISKIE SLĒDZIENI LAUKSAIMNIECĪBAS PĒTĪJUMOS
HYPOTHESIS TESTING AND STATISTICAL INFERENCES IN AGRICULTURAL INVESTIGATIONS

E. Jeņikejeva, I. Arhipova
LLU Informātikas institūts
Institute of Informatics, LUA

Abstract. It is recommended in agricultural investigations to use the ρ value performing statistical inferences. The ρ value for any test is the lowest significance level at which a null hypothesis can be rejected. The (1-ρ) is the maximal probability of acceptance of an alternative hypothesis and the ρ value is the value at which the hypothesis test procedure changes conclusions. The study also considers some examples from cattle breeding.
Key words: hypothesis testing, significance level (type 1 error rate α), power of a tests (type 2 error rate β), ρ -value.

1. Ievads

Hipotēžu pārbaude un statistiskie slēdzieni ir lauksaimniecības pētījumu neatņemama sastāvdaļa. Novērojumu vai eksperimenta datu statistiskās apstrādes rezultātā tiek sastādīti ticamības intervāli ģenerālkopas vidējai vērtībai, salīdzinātas divu izlašu vidējās vērtības un dispersijas, pārbaudīts, vai novērojumu datu izlase normāli sadalīta, noteikta dažādu faktoru ietekme ar dispersijas analīzi, u.t.t. Visu šo problēmu risināšana saistīta ar statistisko hipotēžu pārbaudi.

2. Materiāls un metodes

Hipotēžu pārbaudei ir nepieciešami šādi 5 soļi:
1. solis - formulēt nulles hipotēzi un tai alternatīvo hipotēzi;
2. solis - noteikt atbilstošā testa statistiku (formulu);
3. solis - noteikt hipotēzes noraidīšanas apgabalu (būtiskuma līmeni);
4. solis - aprēķināt testa statistikas vērtību un noraidīt vai nenoraidīt nulles hipotēzi;
5. solis - formulēt slēdzienu reālās problēmas terminos, t.i., apkopot analīzes rezultātus.

Aplūkosim šādu piemēru: pieņemsim, ka mūs interesē, vai būtiski atšķiras divu saimniecību govju izslaukumi laktācijas periodā. Lai atbildētu uz šo jautājumu, mēs sastādījām no katras saimniecības apjoma 50 gadījuma rakstura paraugkopas. Vidējie izslaukumi katrā paraugkopā ir šādi: $\overline{x_1}$ =3880kg, $\overline{x_2}$ =4000kg, dispersijas: s_1^2 =96800 un s_2^2 =95400, $n_1 = n_2$ =50. Pieņemsim, mēs jau pārbaudījām, ka dispersijas statistiski neatšķiras savā starpā (ar varbūtību 95 %). Mēs gribētu apvienot abas paraugkopas vienā, tāpēc mums jāpārbauda, vai atbilstošu ģenerālkopu vidējās vērtības arī statistiski ir vienādas. Izpildīsim šo pārbaudi pa šādiem soļiem (A. H. Kvanli, C. S. Guyness, R. J. Pavur, 1989):

1. solis - H_0: $\mu_1 = \mu_2$ (ģenerālkopu vidējās vērtības ir vienādas);
H_1: $\mu_1 \neq \mu_2$ (ģenerālkopu vidējās vērtības nav vienādas);

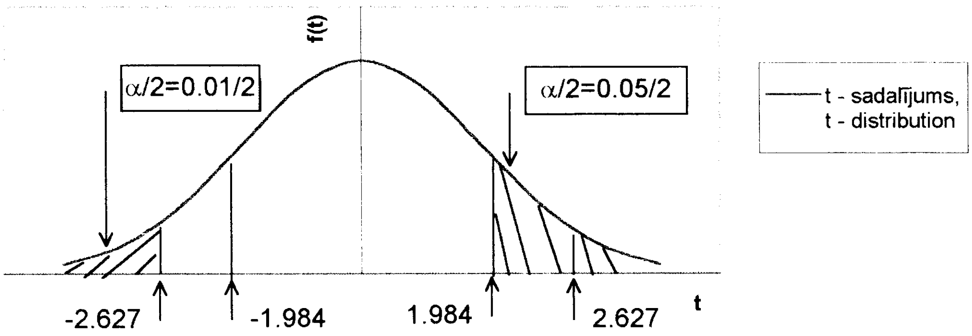
2. solis - pielietosim Stjudenta kritēriju (t-testa statistiku) $t = \frac{|\overline{x_1} - \overline{x_2}|}{\sqrt{s_1^2 \ n_1 + s_2^2 \ n_2}}$ (1);

3. solis - izvēlēsimies būtiskuma līmeni α=0.05 (noraidot H_0 hipotēzi, mēs pieļaujam 5 % kļūdas). Brīvības pakāpju skaits $\nu = n_1 + n_2 - 2$ =98. Kritiskā t vērtība $t_{0.05;98}$ =1.984. H_0 hipotēze jānoraida, ja t > 1.984 vai t < -1.984;

4. solis - aprēķināsim faktisko t vērtību pēc formulas (1): t=1.935. Tad | t | < $t_{0.05;98}$ =1.984 un nevar noraidīt nulles hipotēzi ar varbūtību 95 % (ar būtiskuma līmeni 0.05);

5. solis - nevar secīnāt, ka divu saimniecību govju vidējie izslaukumi būtiski atšķiras (ar varbūtību 95 %).

Tā ir parasta standarta statistisko hipotēžu pārbaudes metode. Svarīgākais šajā metodē ir 3. solis, kurā, fiksējot būtiskuma līmeni α (parasti α=0.05; 0.01; 0.001), faktiski tiek norādīts hipotēzes noraidīšanas (kritiskais) apgabals. Katram būtiskuma līmenim α un brīvības pakāpju skaitam ν atbilst noteiktais kritiskais skaitlis. Piemēram, $t_{0.05;98}$ =1.984; $t_{0.01;98}$=2.627. Ģeometriski būtiskuma līmenim α atbilst laukuma daļa zem sadalījuma blīvuma funkcijas līknes (skat. 1. att.).



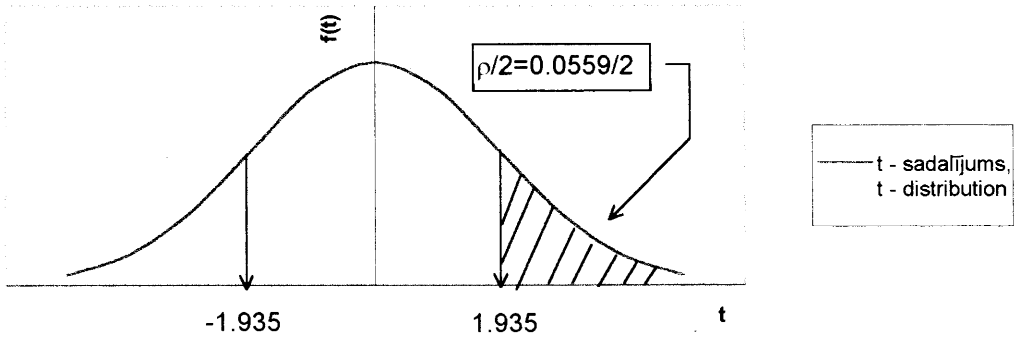
1. att. Kritiskās t-vērtības un atbilstošie kritiskie apgabali.
Fig. 1. Critical t-values and corresponding critical domains.

Mūsu piemēra t-tests ir abpusējs, tāpēc, dalot α ar 2, iegūstam divas "astes", ierobežotas ar sadalījuma blīvuma līkni, abscisu asi un taisnēm t = $t_{\alpha;\nu}$ (labajā astē) un t =-$t_{\alpha;\nu}$ (kreisajā astē). Katras astes laukums ir vienāds ar α/2 un kopējais laukums ir vienāds ar iepriekš uzdoto būtiskuma līmeni α. Abas astes ir nulles hipotēzes pieņemšanas (kritiskais) apgabals. 4. solī aprēķina faktisko t vērtību un salīdzina to ar kritisko $t_{\alpha;\nu}$ vērtību, noraidot (ja t > $t_{\alpha;\nu}$) vai nenoraidot (ja t $\le t_{\alpha;\nu}$) nulles hipotēzi.

Šīs metodes svarīgā vieta: pēc šīs metodes tikai ir fiksēts, vai t pieder hipotēzes noraidīšanas vai hipotēzes pieņemšanas apgabalā, bet netiek ņemts vērā, cik stipri faktiskā t vērtība atšķiras no kritiskās $t_{\alpha;\nu}$ vērtības, kas ir ļoti svarīgi daudzos gadījumos. Salīdzinot aplūkotā piemēra faktisko t vērtību ar kritisko redzam, ka tie atšķiras maz. Būtiskuma līmenim α = 0.06 kritiskā t vērtība $t_{0.06;98}$ =1.903 < t =1.935. Tādējādi, ar varbūtību P=94 % mēs varam noraidīt nulles hipotēzi, t.i., govju vidējie izslaukumi ar varbūtību 94 % atšķiras abās saimniecībās. Šī atbilde ir precīzāka un, galvenais, saturīgāka par 1. atbildi 5. solī. Dotais piemērs rāda, ka, pielietojot standarto hipotēzes parbaudes metodi, rezultāts var izrādīties maznozimīgs.

Lai novērstu šo metodes trūkumu, dažos matemātiskās statistikas darbos (J. E. Hanke, A. G. Retsch, 1990; D. Gujarati, 1992) tiek piedāvāts apgriezt hipotēzes pārbaudes procedūru un nefiksēt apriori būtiskuma līmeni α, bet meklēt kopējo astu laukumu atbilstošai faktiskai t vērtībai. Kopējo astu laukumu sauc par ρ-vērtību (ρ-value). Šī ρ-vērtība ir precīzs būtiskuma līmenis, ko var definēt arī kā mazāko būtiskuma līmeni nulles hipotēzes noraidīšanai.

ρ-vērtibas izmantošana kļuva iespējama sakarā ar datortehnikas attīstīšanu. Mūsdienās daudzās statistiskās paketēs (SAS, MINITAB, STATGRAPHICS, SHAZAM, EXCEL) paredzēta ρ-vērtības aprēķināšana ne tikai t-testam (kā mūsu piemērā), bet arī Z-testam, χ^2- un F-testam. ρ-vērtība ir tāda α vērtība, kurā mainās hipotēzes pārbaudes procedūras secinājums (A. H. Kvanli, C. S. Guyness, R. J. Pavur, 1989). Mūsu piemērā faktiskai t-vērtībai t =1.935 atbilst ρ-vērtība ρ =0.0559 (skat. 2. att.).



2. att. Faktiskā t-vērtība un atbilstošā ρ-vērtība.
Fig. 2. Actual t-value and corresponding ρ-value.

$0.05 < \rho < 0.06$, tāpēc ar varbūtību 95 % mēs nevaram, bet ar varbūtību 94 % - varam noraidīt nulles hipotēzi. Ja ρ-minimālais būtiskuma līmenis, kurā var noraidīt nulles hipotēzi, tad (1-ρ) var interpretēt kā lielāko varbūtību, ar kuru var noraidīt nulles hipotēzi. Pašu ρ vērtību tad var traktēt kā nulles hipotēzes pieņemšanas varbūtību. Mūsu piemērā H_0 hipotēzes noraidīšanas (vai alternatīvās hipotēzes H_1 pieņemšanas) maksimālā varbūtība ir 1- ρ =0.9441, bet nulles hipotēzes piņemšanas varbūtība ir ρ=0.0559. Tādējādi var secināt, ka apskatīto divu saimniecību vidējie izslaukumi atšķiras savā starpā ar varbūtību 0.944 (94.4 %). Ja hipotēzes pārbaudes procedūrā mēs izmantojam ρ-vērtību, hipotēzes pārbaudes nepieciešamie soļi būs šādi:

1. solis - formulēt nulles hipotēzi un tai alternatīvo hipotēzi;
2. solis - noteikt atbilstošā testa statistiku (formulu);
3. solis - aprēķināt testa statistikas vērtību;
4. solis - atrast atbilstošo ρ-vērtību un noraidīt nulles hipotēzi ar varbūtību (1-ρ) (vai nenoraidot to ar varbūtību ρ);
5. solis - formulēt slēdzienu reālās problēmas terminos.

3. Rezultāti

Aplūkosim šādu piemēru. No divām saimniecībām "Druva" un "Vecauce" tika atlasītas divas govju grupas un govju izslaukumi laktācijas periodā. Divu paraugkopu statistiskie rādītāji apkopoti 1. tabulā.

1. tabula/Table 1

Divu paraugkopu statistiskie rādītāji
Two sample statistical measures

Statistiskie rādītāji/paraugkopa Descriptive statistics/sample	"Druva"	"Vecauce"
Vidējā vērtība Average	3660	3585
Dispersija Variance	727014	656639
Paraugkopas apjoms Sample range	60	87

Mēs gribētu apvienot abas paraugkopas vienā. To var izdarīt, ja abām izlasēm ir viens un tas pats sadalījums, un vismaz atbilstošo ģenerālkopu vidējās vērtības būtiski neatšķiras. Aprēķini izpildīti ar EXCEL paketi. T-tests vidējo vērtību salīdzināšanai ir atkarīgs no tā, vai vienādas ir atbilstošo ģenerālkopu dispersijas. To var pārbaudīt ar F-testu, kuru var pielietot, ja atbilstošas ģenerālkopas ir normāli sadalītas. Tāpēc vispirms jāpārbauda, vai katra no paraugkopām ir normāli sadalīta. To izpilda ar χ^2-testu (skat. 2., 3. tab.).

2. tabula/Table 2

Empīriskā ("Vecauce") un normālā sadalījuma atbilstības pārbaude pēc χ^2-testa
Empirical ("Vecauce") and normal distributions and testing the goodness-of-fit with χ^2-test

Intervāla labā robeža The right border of interval	Intervāla vidējā vērtība The average value of interval	Frekvence Frequency	Teorētiskā frekvence (normālais sadalījums) Theoretical frequency (normal distribution)	hi2
2753	2496	13	12.0	0.08
3267	3010	21	17.1	0.88
3781	3524	18	22.0	0.71
4295	4038	17	18.8	0.18
4809	4552	12	10.8	0.13
5837	5580	6	5.2	0.12
Kopā		87	85.9	2.10

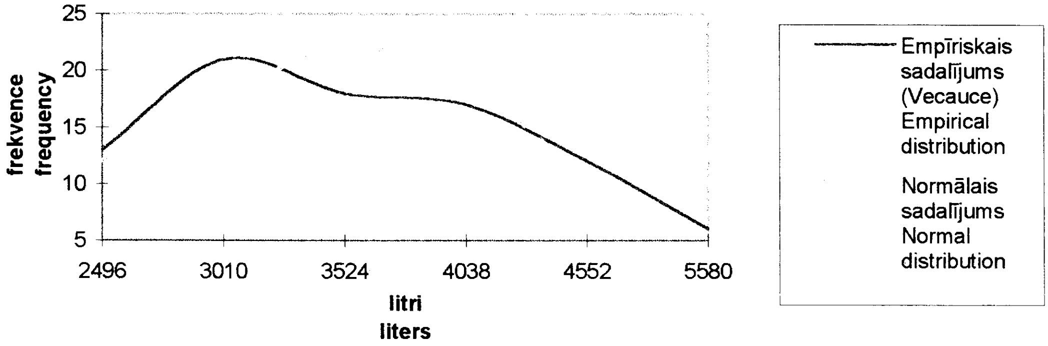
3. tabula/Table 3

Empīriskā ("Druva") un normālā sadalījuma atbilstības pārbaude pēc χ^2-testa
Empirical ("Druva") and normal distribution and testing the goodness-of-fit with χ^2-test

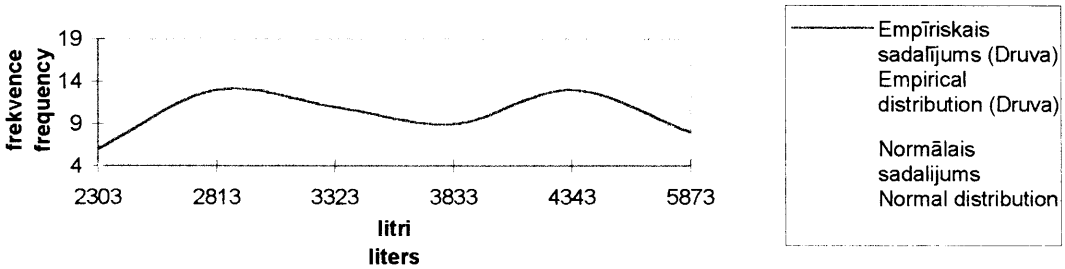
Intervāla labā robeža The right border of interval	Intervāla vidējā vērtība The average value of interval	Frekvence Frequency	Teorētiskā frekvence (normālais sadalījums) Theoretical frequency (normal distribution)	hi2
2558	2303	6	4.0	0.96
3068	2813	13	8.7	2.08
3578	3323	11	13.2	0.38
4088	3833	9	14.0	1.80
4598	4343	13	10.4	0.65
6128	5873	8	7.8	0.01
Kopā		60	58.3	5.88

Faktiskā χ^2-vērtība χ^2=2.10, bet kritiskā $\chi^2_{0.05,3}$=7.81, tad ar varbūtību 95 % mēs nevaram noraidīt, ka sadalījums ir normālais. Parasti tiek pieņemts nepareizs secinājums, ka empīriskais sadalījums atbilst normālajam. Bet faktiskai χ^2-vērtībai χ^2=2.10 atbilst ρ-vērtība ρ=0.5512. Tāpēc ar varbūtību 1-ρ =0.4498 varam noraidīt nulles hipotēzi un pieņemt, ka sadalījums neatbilst normālam ar varbūtību 0.4498 (skat. 3. att.)

Faktiskā χ^2-vērtība χ^2=5.88, bet kritiskā $\chi^2_{0.05,3}$=7.81, tad ar varbūtību 95 % mēs nevaram noraidīt, ka sadalījums ir normālais. Parasti no tā pieņem nepareizu secinājumu, ka empīriskais sadalījums atbilst normālajam. Bet faktiskai χ^2-vērtībai χ^2=5.88 atbilst ρ-vērtība ρ=0.1176. Tāpēc ar varbūtību 1-ρ =0.8824 varam noraidīt nulles hipotēzi un pieņemt, ka sadalījums neatbilst normālam ar varbūtību 0.8824 (skat. 4. att.).



3. att. Empīriskā ("Vecauce") un normālā sadalījuma līknes.
Fig. 3. Empirical ("Vecauce") and normal distributions.



4. att. Empīriskā ("Druva") un normālā sadalījuma līknes.
Fig. 4. Empirical ("Druva") and normal distributions.

Šie aprēķini rāda, ka "Vecauces" paraugkopa atšķiras no normālā sadalījuma par 45 %, bet "Druvas" paraugkopa atšķiras no normālā par 88 %. No pēdējā grafika redzams, ka "Druvas" paraugkopa vispār ir bimodālā, kas liecina par to, ka šī paraugkopa nav homogēna, tajā ir pārstāvētas divas paraugkopas. Pirms "Druvas" paraugkopas statistiskas apstrādes un analīzes ir nepieciešami papildnovērojumi un vairāk pārdomāta novērojumu izlases sistēma.

4. Slēdziens

Noslēgumā atzīmēsim, ka ρ vērtība jāatšķir no *kritērija jaudas 1-β*. Būtiskuma līmenis α un kritērija jauda 1-β tika sīki apskatīti J. Neimana un Pirsona darbos 30-os gados (J. Neyman, E. S. Pearson, 1933). Šajos darbos vispirms tiek noteikts fiksētais būtiskuma līmenis α un tikai pēc tam fiksētam α tiek atrasta (fiksēta) β-nepareizās hipotēzas pieņemšanas kļūda. Kā atzīmējis G. Kramers (Г. Крамер,1975), dotajam būtiskuma līmenim α kritērija jauda tiek noteikta neviennozīmīgi. Svarīgākā problēma - salīdzināt dažādus kritērijus dotajai hipotēzei un izvēlēties labāko. Šī problēma ir ļoti grūta. Vienkāršākajos gadījumos to atrisinājumu var atrast klasiskajos Pirsona, Neimana un Kramera darbos (J. Neyman, E. S. Pearson, 1933; Г. Крамер, 1975). Bet mūsu uzdevums šajā rakstā ir vienkaršāks - katrā atsevišķā gadījumā (katrai dotā kritērija vērtībai) noteikt ρ vērtību, lai ar šo varbūtību nenoraidītu (ja ρ ir liels) vai ar varbūtību (1-ρ) noraidītu nulles hipotēzi.

Literatūra

1. Alan H. Kvanli, C. Stepfen Guyness, Robert J. Pavur. (1989). Introduction to Business Statistics. A computer integrated approach. University of North Texas. 899.
2. Damodar Gujarati. (1992). Essentials of Econometrics.McGRAW-HILL, INC. 466.
3. John E. Hanke, Arthur G. Retsch. (1990). Understanding Business Statistics. United States of America. 871.
4. Neyman J. and Pearson E. S. (1933). On the problem of the most efficient tests of statistical hypothesis. PTRS. 289.
5. Крамер Г. (1975). Математические методы статистики. Москва. 648.